SAM-TDR-62-149

403 917

63 3

(5) 787 750

AD No. 403917

ASTIA FILE COPY

ANALYSIS OF REPEATED MEASUREMENTS WITH DISPROPORTIONATE SAMPLES FOR TREATMENTS IN A TWO-WAY CLASSIFICATION

(7) NA
(8) NA
(9) NA
(10) [illegible]
(11) [illegible] 62,
(12) 7p.
(13-17) NA
(18) SAM/B
(19) TDR 62-149

TECHNICAL DOCUMENTARY REPORT NO. SAM-TDR-62-149

December 1962

USAF School of Aerospace Medicine
Aerospace Medical Division (AFSC)
Brooks Air Force Base, Texas

Task No. 775505

$1.10

FOREWORD

This report was prepared by the following personnel in the Department of Biometrics:

(1D) RICHARD C. McNEE, ~~B.S.~~ by
PHELPS P. CRUMP, ~~Ph.D.~~ and

ABSTRACT

A procedure is given for computing the analysis of variance obtained by fitting constants of the experimental setup where repeated measurements are made on subjects over time and treatments have a two-way classification with disproportionate subclass numbers. Under certain assumptions, this analysis gives a test of the three-factor interaction, tests of two-factor interactions if there is no three-factor interaction, and tests of main effects if there are no interactions. The advantage of this procedure over the standard procedure of fitting constants is in the amount of computation involved.

This technical documentary report has been reviewed and is approved.

Robert B. Payne

ROBERT B. PAYNE
Colonel, USAF, MSC
Chief, Operations Division

ANALYSIS OF REPEATED MEASUREMENTS WITH DISPROPORTIONATE SAMPLES FOR TREATMENTS IN A TWO-WAY CLASSIFICATION

1. INTRODUCTION

In the analysis of data from a repeated measurement design in which the *treatments* are arranged in a two-way classification and the subclass numbers are disproportionate, difficulties arise similar in nature to those encountered in the analysis of data from a simple two-way classification experiment with disproportionate subclass numbers. In the present paper the computational procedures for a method of analysis for this setup are given along with a numerical example. The theoretic development and justification for the proposed method will be presented in a later paper.

In studying the effects of *treatments* on subjects in time, it is sometimes desirable to design a repeated measurements experiment with a two-way classification of *treatments.* For example, a physiologist may be interested in studying the effect over time of certain drugs at different altitudes. It would be desirable to have the drugs and altitudes in a factorial arrangement so as to be able to determine whether or not the relative effects of the drugs are the same at all altitudes. As another example, the experimenter may be interested in whether males and females respond the same to different levels of a treatment. In this case sex becomes the second way of classification.

Disproportionality of the numbers of experimental units in such setups can occur for many reasons. It can be present from the beginning in experiments when a sufficient number of appropriate experimental units is not available. Further, even if an experiment is initially set up with proportionate (which includes equal) allocations, the experimenter frequently ends up with disproportionate numbers for a variety of reasons. For example, subjects may be dropped from the experiment owing to a sickness which is completely unrelated to the effects of the treatments, or part of the experimental material may be accidentally lost. When subjects are not proportionately allocated, the computation of the sums of squares in the analysis of variance is determined by which, if any, of the interactions can be assumed zero, just as in the case of the two-way classification experiment with disproportionate subclass numbers.

Under the symmetry assumption, which will be explained below, the procedure presented herein will give a test for the three-factor interaction. If the three-factor interaction is zero, it then provides tests for the two-factor interactions. Finally, if all interactions are zero, it provides tests for the main effects. This analysis is an extension of the method for the two-way classification frequently referred to as the "Method of Fitting Constants" (1).

The *symmetry* assumption states that the observations for each subject are equally correlated in time and the variances at all *times* are equal. This assumption makes

Received for publication on 9 October 1962.

the study of the distributions for the usual test statistics analytically tractable. See Hughes and Danford (2) for an exposition of this requirement for the repeated measurements case when *treatments* have a single classification. When the symmetry assumption cannot be made the only method of complete analysis available at the present for the repeated measurements design with singly classified treatments is a multivariate procedure (3). The procedure proposed herein is valid only for the symmetry assumption.

2. MODEL EQUATION

For the experimental setup considered in this paper, the model equation is:

$$Y_{ijkm} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \gamma_k + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk} + s_{ijm} + g_{ijkm}.$$

In the equation given above, Greek letters stand for fixed effects and Latin letters for random effects, with:

μ = the overall mean

α_i = the effect of the i^{th} level of factor A ($i = 1, \ldots r$)

β_j = the effect of the j^{th} level of factor B ($j = 1, \ldots q$)

γ_k = the effect of the k^{th} time ($k = 1, \ldots p$)

$\alpha\beta_{ij}$ = the interaction effect of the i^{th} level of factor A and the j^{th} level of factor B

$\alpha\gamma_{ik}$ = the interaction effect of the i^{th} level of factor A and the k^{th} time

$\beta\gamma_{jk}$ = the interaction effect of the j^{th} level of factor B and the k^{th} time

$\alpha\beta\gamma_{ijk}$ = the interaction effect of the i^{th} level of factor A, the j^{th} level of factor B, and the k^{th} time

s_{ijm} = the effect of the m^{th} subject in the i, j^{th} group ($m = 1, \ldots n_{ij}$) with the assumptions:

$$E(s_{ijm}) = 0$$
$$E(s_{ijm}, s_{i'j'm'}) = \begin{cases} \sigma_s^2; & i = i', j = j', m = m \\ 0; & \text{otherwise} \end{cases}$$

g_{ijkm} = the random error term plus the inseparable (for this design) interaction effect of the m^{th} subject in the i, j^{th} group and the k^{th} time, with the assumptions:

$$E(g_{ijkm}) = 0$$
$$E(g_{ijkm}, g_{i'j'k'm'}) = \begin{cases} \sigma^2; & i = i', j = j', k = k', m = m' \\ 0; & \text{otherwise} \end{cases}$$

Finally, for purposes of testing later on, it is assumed that the s_{ijm} and g_{ijkm} are *normally* distributed and the s_{ijm} are independent of the g_{ijkm} for all i, j, k, m. Writing the model with a random subject effect and uncorrelated *error* implies that the observations for a subject are equally correlated in time. This, with the assumption that the variances at all times are equal, which is implied in the assumptions on the s_{ijm} and g_{ijkm}, is the symmetry assumption.

3. ANALYSIS

Certain sums of squares are unaffected by the disproportionality and take the usual form. These are:

$$\text{SS (time)} = \sum_k Y^2_{..k.}/n.. - Y^2_{....}/pn..$$

SS (Error (a)) = SS (Subjects/groups)

$$= \sum_{i}\sum_{j}\sum_{m} Y^2_{ij\cdot m}/p - \sum_{i}\sum_{j} Y^2_{ij\cdot\cdot}/pn_{ij}$$

SS (Error (b)) = SS (Subject $\times$ time/groups)

$$= \sum_{i}\sum_{j}\sum_{k}\sum_{m} Y^2_{ijkm} - \sum_{i}\sum_{j}\sum_{k} Y^2_{ijk\cdot}/n_{ij} - \text{SS (Error (a))}$$

where $n_{\cdot\cdot} = \sum_{i}\sum_{j} n_{ij}$ and $Y_{\cdot\cdot k\cdot} = \sum_{i}\sum_{j}\sum_{m} Y_{ijkm}$; i.e., the dot indicates summation over the index it replaces. The other sums of squares required in the analysis of variance can be most clearly shown by using matrix notation in part. For that purpose the quantities C and G_k^T will be utilized, where $r' = r - 1$, $q' = q - 1$, and G_k^T is the transpose of G_k. Using this notation, the sums of squares in the analysis of variance can be computed as shown in table I, where again the dot indicates summation over the index it replaces; e.g., $G_\cdot^T = \sum_k G_k^T$. Thus, in order to compute the sums of squares it is only necessary to calculate certain standard sums of squares, obtain the inverse of a matrix of order $r + q - 1$, and perform some simple matrix multiplications. **The above procedure for computing sums of squares is equivalent to fitting constants, but affords simpler computational forms.**

$$C = \begin{bmatrix}
n_{\cdot\cdot} & n_{1\cdot} & n_{2\cdot} & \cdots & n_{r'\cdot} & n_{\cdot 1} & n_{\cdot 2} & \cdots & n_{\cdot q'} \\
n_{1\cdot} & n_{1\cdot} & 0 & \cdots & 0 & n_{11} & n_{12} & \cdots & n_{1q'} \\
n_{2\cdot} & 0 & n_{2\cdot} & \ddots & \vdots & n_{21} & n_{22} & \cdots & n_{2q'} \\
\vdots & \vdots & \ddots & \ddots & 0 & \vdots & \vdots & & \vdots \\
n_{r'\cdot} & 0 & \cdots & 0 & n_{r'\cdot} & n_{r'1} & n_{r'2} & \cdots & n_{r'q'} \\
n_{\cdot 1} & n_{11} & n_{21} & \cdots & n_{r'1} & n_{\cdot 1} & 0 & \cdots & 0 \\
n_{\cdot 2} & n_{12} & n_{22} & \cdots & n_{r'2} & 0 & n_{\cdot 2} & \ddots & \vdots \\
\vdots & \vdots & \vdots & & \vdots & \vdots & \ddots & \ddots & 0 \\
n_{\cdot q'} & n_{1q'} & n_{2q'} & \cdots & n_{r'q'} & 0 & \cdots & 0 & n_{\cdot q'}
\end{bmatrix}$$

$$G_k^T = [Y_{\cdot\cdot k\cdot},\ Y_{1\cdot k\cdot},\ Y_{2\cdot k\cdot}, \cdots\ Y_{r'\cdot k\cdot},\ Y_{\cdot 1k\cdot},\ Y_{\cdot 2k\cdot}, \cdots\ Y_{\cdot q'k\cdot}]$$

TABLE I

Sums of squares and degrees of freedom for the analysis of variance

Source	d.f.	S. Sq.
A	$r-1$	$G^{T}_{.}C^{-1}G_{.}/p - \sum_{j} Y^{2}_{.j..}/pn_{.j} = (1)$
B	$q-1$	$G^{T}_{.}C^{-1}G_{.}/p - \sum_{i} Y^{2}_{i...}/pn_{i.} = (2)$
A $\times$ B	$(r-1)(q-1)$	$\sum_{i}\sum_{j} Y^{2}_{ij..}/pn_{ij} - G^{T}_{.}C^{-1}G_{.}/p = (3)$
Error (a)	$n_{..} - rq$	$\sum_{i}\sum_{j}\sum_{m} Y^{2}_{ij.m}/p - \sum_{i}\sum_{j} Y^{2}_{ij..}/pn_{ij} = (4)$
Time	$p-1$	$\sum_{k} Y^{2}_{..k.}/n_{..} - Y^{2}_{....}/pn_{..}$
A $\times$ t	$(r-1)(p-1)$	$\sum_{k} G^{T}_{k}C^{-1}G_{k} - \sum_{j}\sum_{k} Y^{2}_{.jk.}/n_{.j} - (1)$
B $\times$ t	$(q-1)(p-1)$	$\sum_{k} G^{T}_{k}C^{-1}G_{k} - \sum_{i}\sum_{k} Y^{2}_{i.k.}/n_{i.} - (2)$
A $\times$ B $\times$ t	$(r-1)(q-1)(p-1)$	$\sum_{i}\sum_{j}\sum_{k} Y^{2}_{ijk.}/n_{ij} - \sum_{k} G^{T}_{k}C^{-1}G_{k} - (3)$
Error (b)	$(n_{..} - rq)(p-1)$	$\sum_{i}\sum_{j}\sum_{k}\sum_{m} Y^{2}_{ijkm} - \sum_{i}\sum_{j}\sum_{k} Y^{2}_{ijk.}/n_{ij} - (4)$
Total	$pn_{..} - 1$	*

*Not given since individual sums of squares do not sum to total sum of squares.

Under the assumptions given in the model statement, the Error (a) mean square is the appropriate mean square for testing the effects "above the line" and Error (b) mean square is used for testing the effects "below the line." In this analysis, the sum of squares for the three-factor interaction is adjusted for all two-factor interactions and main effects, the sum of squares for each two-factor interaction is adjusted for all other two-factor interactions and all main effects, and the sum of squares for each main effect is adjusted for all other main effects. However, the sums of squares for the two-factor interactions are not adjusted for the three-factor interaction and the sums of squares for the main effects are not adjusted for any interactions. Therefore, if the three-factor interaction is significant, the main effects and two-factor interactions should be looked at in two-way tables at each level of the third factor and if there are significant first-order interactions, the main effects of the factors in the significant interaction should be looked at separately at each level of the other factor. If overall tests of significance for main effects or two-factor interactions are desired in the above-named cases, then a different analysis should be performed.

4. EXAMPLE

The basic data for the example are given in table II. There are three levels of factor A, two levels of factor B, and three measurements made over time, with the time points spaced the same for all subjects. The number of subjects for each treatment combination, n_{ij}, is given in table III. These are constant over time; i.e., $p = 3$ observations were made over time on each subject.

TABLE II

Basic data for the example

Treatment A	Treatment B	Time 1	Time 2	Time 3
1	1	39	36	38
		38	42	30
		40	28	33
	2	36	39	38
		31	34	28
		46	44	35
		44	41	36
		32	34	33
2	1	56	51	53
		54	40	45
		50	45	49
		32	33	31
	2	39	34	32
		54	56	50
3	1	38	41	37
		39	33	40
		51	38	39
	2	55	48	46
		39	44	30
		49	36	42
		44	39	22

TABLE III

The number of subjects with each treatment combination for the example

Treatment A	Treatment B 1	Treatment B 2	Total
1	3	5	8
2	4	2	6
3	3	4	7
Total	10	11	21

The C matrix is determined from table III:

$$C = \begin{bmatrix} n_{\cdot\cdot} & n_{1\cdot} & n_{2\cdot} & n_{\cdot 1} \\ n_{1\cdot} & n_{1\cdot} & 0 & n_{11} \\ n_{2\cdot} & 0 & n_{2\cdot} & n_{21} \\ n_{\cdot 1} & n_{11} & n_{21} & n_{\cdot 1} \end{bmatrix} = \begin{bmatrix} 21 & 8 & 6 & 10 \\ 8 & 8 & 0 & 3 \\ 6 & 0 & 6 & 4 \\ 10 & 3 & 4 & 10 \end{bmatrix}$$

Its inverse is:

$$C^{-1} = \begin{bmatrix} .180169286 & -.147521161 & -.122128174 & -.087061669 \\ -.147521161 & .268440145 & .140266022 & .010882709 \\ -.122128174 & .140266022 & .321039903 & -.048367593 \\ -.087061669 & .010882709 & -.048367593 & .203143894 \end{bmatrix}$$

The G_k are determined from table II:

$$G_1 = \begin{bmatrix} 906 \\ 306 \\ 285 \\ 437 \end{bmatrix} ; \quad G_2 = \begin{bmatrix} 836 \\ 298 \\ 259 \\ 387 \end{bmatrix} ; \quad G_3 = \begin{bmatrix} 787 \\ 271 \\ 260 \\ 395 \end{bmatrix}$$

From these quantities, calculate:

$$G_1^T C^{-1} G_1 = 39{,}418.54$$
$$G_2^T C^{-1} G_2 = 33{,}459.44$$
$$G_3^T C^{-1} G_3 = 29{,}830.74$$
$$G_{\cdot}^T C^{-1} G_{\cdot} / p = 102{,}223.80$$
$$\sum_k G_k^T C^{-1} G_k = 102{,}708.72$$

Furthermore:

$$\sum_i Y_{i\cdots}^2 / pn_{i\cdot} = 102{,}217.80$$

$$\sum_j Y_{\cdot j\cdot\cdot}^2 / pn_{\cdot j} = 101{,}535.06$$

$$\sum_i \sum_j Y_{ij\cdot\cdot}^2 / pn_{ij} = 102{,}236.43$$

$$\sum_k Y_{\cdot\cdot k\cdot}^2 / n_{\cdot\cdot} = 101{,}861.95$$

$$\sum_i \sum_k Y_{i\cdot k\cdot}^2 / n_{i\cdot} = 102{,}626.89$$

$$\sum_j \sum_k Y_{\cdot jk\cdot}^2 / n_{\cdot j} = 101{,}969.57$$

$$\sum_i \sum_j \sum_k Y_{ijk\cdot}^2 / n_{ij} = 102{,}761.92$$

$$\sum_i \sum_j \sum_m Y^2_{ij\cdot m}/p = 104{,}187.00$$

$$\sum_i \sum_j \sum_k \sum_m Y^2_{ijkm} = 105{,}277.00$$

$$Y^2_{\cdots\cdot}/pn_{\cdot\cdot} = 101{,}521.29$$

The completed analysis of variance for the example is given in table IV.

TABLE IV

Analysis of variance for the example

Source	d.f.	S. Sq.	M. Sq.	F	P
A	2	688.74	344.37	2.65	NS
B	1	6.00	6.00	.05	NS
A × B	2	12.63	6.31	.05	NS
Error (a)	15	1,950.57	130.04		
Time	2	340.66	170.33	9.05	< .001
A × t	4	50.41	12.60	.67	NS
B × t	2	75.83	37.92	2.01	NS
A × B × t	4	40.57	10.14	.54	NS
Error (b)	30	564.51	18.82		
Total	62				

In the example, all of the probability levels for the interactions are greater than .10. Therefore, it seems reasonable to assume that all of the interactions are zero. (It would seem desirable to use a probability level greater than .05 to test this assumption.) Under this assumption, the F-ratios for main effects give valid tests of the main effects. That is, the F-ratios give tests of main effects only, rather than main effects plus interactions as they would if interactions were not zero. From this analysis, there is no indication of an effect due to either factor A or factor B. However, there is a highly significant time effect.

Even in an experiment as small as this example, the advantage of this procedure is apparent. One of the several matrices to be inverted using the standard procedure of fitting constants would be of order 12, whereas with this procedure only one matrix of order 4 need be inverted.

REFERENCES

1. Snedecor, G. W. Statistical methods, 5th ed. Ames, Iowa: Iowa State College Press, 1956.
2. Hughes, H. M., and M. B. Danford. Repeated measurement designs, assuming equal variances and covariances. School of Aviation Medicine Report 59-40, Dec. 1958.
3. Danford, M. B., H. M. Hughes, and R. C. McNee. On the analysis of repeated measurements experiments. Biometrics 16:547-565 (1960).

USAF School of Aerospace Medicine, Brooks AF Base, Tex.

SAM-TDR-62-149. ANALYSIS OF REPEATED MEASUREMENTS WITH DISPROPORTIONATE SAMPLES FOR TREATMENTS IN A TWO-WAY CLASSIFICATION. Dec. 62, 7 pp. incl. tables, 3 refs.

Unclassified Report

A procedure is given for computing the analysis of variance obtained by fitting constants of the experimental setup where repeated measurements are made on subjects over time and treatments have a two-way classification with disproportionate subclass

1. Statistical analysis
2. Analysis of variance
3. Univariate analysis
4. Repeated measurements design
5. Disproportionate subclass numbers

I. AFSC Task 775505
II. McNee, R. C., Crump, P. P.
III. In ASTIA collection

USAF School of Aerospace Medicine, Brooks AF Base, Tex.

SAM-TDR-62-149. ANALYSIS OF REPEATED MEASUREMENTS WITH DISPROPORTIONATE SAMPLES FOR TREATMENTS IN A TWO-WAY CLASSIFICATION. Dec. 62, 7 pp. incl. tables, 3 refs.

Unclassified Report

A procedure is given for computing the analysis of variance obtained by fitting constants of the experimental setup where repeated measurements are made on subjects over time and treatments have a two-way classification with disproportionate subclass

1. Statistical analysis
2. Analysis of variance
3. Univariate analysis
4. Repeated measurements design
5. Disproportionate subclass numbers

I. AFSC Task 775505
II. McNee, R. C., Crump, P. P.
III. In ASTIA collection

USAF School of Aerospace Medicine, Brooks AF Base, Tex.

SAM-TDR-62-149. ANALYSIS OF REPEATED MEASUREMENTS WITH DISPROPORTIONATE SAMPLES FOR TREATMENTS IN A TWO-WAY CLASSIFICATION. Dec. 62, 7 pp. incl. tables, 3 refs.

Unclassified Report

A procedure is given for computing the analysis of variance obtained by fitting constants of the experimental setup where repeated measurements are made on subjects over time and treatments have a two-way classification with disproportionate subclass

1. Statistical analysis
2. Analysis of variance
3. Univariate analysis
4. Repeated measurements design
5. Disproportionate subclass numbers

I. AFSC Task 775505
II. McNee, R. C., Crump, P. P.
III. In ASTIA collection

USAF School of Aerospace Medicine, Brooks AF Base, Tex.

SAM-TDR-62-149. ANALYSIS OF REPEATED MEASUREMENTS WITH DISPROPORTIONATE SAMPLES FOR TREATMENTS IN A TWO-WAY CLASSIFICATION. Dec. 62, 7 pp. incl. tables, 3 refs.

Unclassified Report

A procedure is given for computing the analysis of variance obtained by fitting constants of the experimental setup where repeated measurements are made on subjects over time and treatments have a two-way classification with disproportionate subclass

1. Statistical analysis
2. Analysis of variance
3. Univariate analysis
4. Repeated measurements design
5. Disproportionate subclass numbers

I. AFSC Task 775505
II. McNee, R. C., Crump, P. P.
III. In ASTIA collection

numbers. Under certain assumptions, this analysis gives a test of the three-factor interaction, tests of two-factor interactions if there is no three-factor interaction, and tests of main effects if there are no interactions. The advantage of this procedure over the standard procedure of fitting constants is in the amount of computation involved.

numbers. Under certain assumptions, this analysis gives a test of the three-factor interaction, tests of two-factor interactions if there is no three-factor interaction, and tests of main effects if there are no interactions. The advantage of this procedure over the standard procedure of fitting constants is in the amount of computation involved.

numbers. Under certain assumptions, this analysis gives a test of the three-factor interaction, tests of two-factor interactions if there is no three-factor interaction, and tests of main effects if there are no interactions. The advantage of this procedure over the standard procedure of fitting constants is in the amount of computation involved.

numbers. Under certain assumptions, this analysis gives a test of the three-factor interaction, tests of two-factor interactions if there is no three-factor interaction, and tests of main effects if there are no interactions. The advantage of this procedure over the standard procedure of fitting constants is in the amount of computation involved.